

# Repository Interoperability with the Hub and Spoke

**Thomas G. Habing, Timothy W. Cole, and William H. Mischo**  
**University of Illinois at Urbana-Champaign**

The University of Illinois at Urbana-Champaign's National Digital Information Infrastructure Preservation Program (NDIIPP) project developed a repository test bed consisting of installations of DSpace, Fedora, EPrints, and Greenstone. During the development of this test bed as we ingested and moved data between repositories we realized that an interoperability layer would be useful, not only to aid our evaluations, but also for the wider community.

With the focus of our project being preservation we felt that being able to easily move digital packages between repositories would facilitate the long-term preservation of those objects. In addition, we also realized from our own situation at UIUC and also from talking with peer institutions that it is common for an institution to have multiple repositories. For example, at UIUC we have a deployment of DSpace as an institutional repository, we are using CONTENTdm for collections of digitized images, we are using DLXS from Michigan for other collections, plus we are investigating Fedora for an upcoming project. This is just within the Library; we know that across the University other departments have digital repositories. We feel this is currently a typical situation, and we also know that these heterogeneous silos of data can create challenges to a unified preservation strategy.

The OAI-PMH has done much to facilitate the wide-spread sharing of metadata, but there are no similar standards that do the same for digital objects. Therefore, we undertook the development of a repository interoperability architecture plus a proof-of-concept implementation which we are calling the "Hub and Spoke (HaS)."

We are designing the HaS not only to provide basic interoperability between repositories, but we are looking forward to using the architecture as a unifying platform for preservation. The

architecture currently consists of several parts. Probably the most important being the METS profile which remains mostly neutral regarding content files and structure but defines a minimum level of descriptive (MODS) and administrative (PREMIS) metadata with a strong emphasis on preservation, particularly technical data and provenance.

A second key part of the architecture is a RESTful Web Service which we call the Lightweight Repository CRUD Service (LRCRUDS). CRUD being Create, Retrieve, Update, and Delete from the relational database lexicon. This service would map URIs to local identifiers and use the HTTP methods PUT, GET, POST, and DELETE to handle packages being submitted or disseminated from a particular repository. Packages would be shipped over the network as Zip archives containing a standard header, METS file, and content files in a format suitable for import into the repository. An instance of this service would be created for each repository participating in the HnS.

The third critical part of the architecture is the Hub. The Hub performs the transformations required to convert from and to the METS profile, and it manages the generation and validation of technical and provenance metadata. Initially the Hub is simply a staging area as packages are moved about. However, the ultimate goal is to develop the Hub into a trusted preservation archive capable of handling submissions from various other repositories and also disseminating packages to those repositories, as well as handling all the other required functions of a trusted digital repository.